Fast Efficient Importance Sampling by State Space Methods

Koopman, S.J.; Nguyen, T.M.

2012

TI 2012-008/4
Tinbergen Institute Discussion Paper

Fast Efficient Importance Sampling by State Space Methods

Siem Jan Koopman[1]
Thuy Minh Nguyen[2]

[1] *Faculty of Economics and Business Administration, VU University Amsterdam, and Tinbergen Institute;*
[2] *Deutsche Bank London.*

Fast Efficient Importance Sampling by State Space Methods

Siem Jan Koopman[(a,b)] *Thuy Minh Nguyen*[(c)]

[(a)] Department of Econometrics, VU University Amsterdam

[(b)] Tinbergen Institute

[(c)] Deutsche Bank, London

January 30, 2012

Some keywords: Kalman filter; Monte Carlo maximum likelihood; Simulation smoothing.
JEL classification: C32, C51.

Acknowledgements: We would like to thank André Lucas and Marcel Scharth for giving comments on an earlier version of this paper. The usual disclaimer applies.

Address of correspondence: S.J. Koopman, Department of Econometrics, VU University Amsterdam, De Boelelaan 1105, NL-1081 HV Amsterdam, The Netherlands.
Email: `s.j.koopman@vu.nl`

Fast Efficient Importance Sampling by State Space Methods

Siem Jan Koopman and Thuy Minh Nguyen

Abstract

We show that efficient importance sampling for nonlinear non-Gaussian state space models can be implemented by computationally efficient Kalman filter and smoothing methods. The result provides some new insights but it primarily leads to a simple and fast method for efficient importance sampling. A simulation study and empirical illustration provide some evidence of the computational gains.

1 Introduction

For the modeling of an observed time series $y_1, \ldots, y_n$, we consider a parametric model that we formulate conditionally on a time-varying parameter vector θ_t. The conditional model for the observations is given by

$$y_t|\theta_t \sim p(y_t|\theta_t; \psi), \qquad t = 1, \ldots, n, \tag{1}$$

where ψ is a fixed parameter vector, with observation density $p(y_t|\theta_t; \psi)$ that is possibly non-Gaussian and may represent a nonlinear relation between y_t and θ_t. Conditional on $\theta_1, \ldots, \theta_n$, the observations $y_1, \ldots, y_n$ are serially independent. The time-varying parameters in θ_t can represent different features of the model including constant, variance and regression coefficients. Different dynamic specifications for the parameters in θ_t can be adopted. In our analysis, the conditional observation density and the dynamic model for θ_t must be specified but both may depend on the fixed parameter vector ψ.

When (i) the conditional observation density is Gaussian, (ii) the relation between y_t and θ_t is linear and (iii) the dynamic model for θ_t is linear Gaussian, our time series modelling framework reduces to the linear Gaussian state space model as discussed, for example, in Durbin and Koopman (2001, Part I). In this framework, we can rely on the celebrated Kalman filter and its related smoothing method for the signal extraction of θ_t, the maximum likelihood estimation of ψ and the forecasting of y_t. However, when we depart from one of the three assumptions, these analyses cannot be carried out using the Kalman filter. We then rely on numerical methods to evaluate the high-dimensional integrals that are required for such analyses. In this paper we discuss Monte Carlo estimation based on importance sampling methods.

The general ideas of importance sampling are established in statistics and econometrics, see Kloek and Van Dijk (1978), Ripley (1987), and Geweke (1989). Importance sampling techniques for state space models have been explored by Danielsson and Richard (1993), Shephard and Pitt (1997), Durbin and Koopman (1997), So (2003) and Jungbacker and Koopman (2007). A textbook treatment is given by Durbin and Koopman (2001). The performance of this approach to Monte Carlo estimation relies on the successful construction of an importance density. Several methods for designing an importance density for time series modelling have been proposed. For example, Shephard and Pitt (1997) and Durbin and Koopman (1997) adopt an importance density based on the mode of the conditional density.

In this paper we consider the efficient importance sampling (EIS) method of Liesenfeld and Richard (2003) and Richard and Zhang (2007) where the sampling is based on a global approximation of the original model. We show that EIS can be implemented using standard Kalman filter methods. It leads to a simple and fast procedure for importance sampling. We discuss how the resulting EIS procedure is related to the procedure of Shephard and Pitt (1997) and Durbin and Koopman (1997), hereafter referred to as SPDK. A simulation study provides the evidence of the computational efficiency gains.

The remainder of the paper is organized as follows. In Section 2 we review the EIS and SPDK methods for constructing the importance density for nonlinear non-Gaussian state space models. In Section 3 we show how the EIS method can be implemented using state space methods and illustrate the gains of our modified efficient importance sampling method by a simulation study. In Section 4, details are provided for an effective implementation of the method. In Section 5, the method is applied to a time-varying model for counts in a simulation study and a stochastic volatility model for daily returns of pound/dollar exchange rates in an empirical study.

2 The model

The dynamic model specification under consideration is given by the observation density $p(y_t|\theta_t;\psi)$ as introduced in (1) and with the stochastically time-varying parameter vector θ_t specified as

$$\theta_t = Z_t(\alpha_t), \qquad t = 1, \ldots, n, \tag{2}$$

where $Z_t(\cdot)$ is a fixed known function that may depend on the parameter vector ψ and where α_t is the stochastically time-varying state vector. The linear Gaussian dynamic process for

α_t is given by

$$\alpha_{t+1} = T_t\alpha_t + R_t\eta_t, \qquad \eta_t \sim \mathrm{NID}(0, Q_t), \qquad \alpha_1 \sim \mathrm{N}(0, Q_0), \tag{3}$$

where the elements of the matrices T_t and R_t and the variance matrices Q_t and Q_0 are known except that some elements have a possible dependence on parameter vector ψ, for $t = 1, \ldots, n$. The disturbances η_t are normally independently distributed, serially uncorrelated and are not dependent on the normally distributed initial state vector α_1. All stochastic and non-stochastic variables have appropriate dimensions and they will be given only when it is necessary. The observation y_t is typically a scalar but the methods presented in Section 3 are also applicable for a vector of observations y_t. Illustrations of special cases of our general modelling framework are given below.

Signal plus heavy-tailed error model

When the time series observations y_t are randomly contaminated by noise with with large shocks, we may wish to remove the noise from the signal and to model the noise explicitly by a heavy-tailed density. We then may consider the model

$$y_t = \theta_t + \varepsilon_t, \qquad \varepsilon_t \sim \tau(0, \sigma^2, \nu), \qquad t = 1, \ldots, n, \tag{4}$$

where $\tau(\mu, \sigma^2, \nu)$ refers to the Student's t density with mean μ, variance σ^2 and degrees of freedom ν. The dynamic specification for θ_t can be determined by (2) and (3). The model clearly fits in our general framework with $p(y_t|\theta_t; \psi) = \tau(\theta_t, \sigma^2, \nu)$ in (1).

Stochastic volatility model

A time series of financial returns is mostly subject to clusters of volatility changes which can effectively be modelled by a dynamic processes for the variance. A basic version of the stochastic volatility model for a time series of returns y_t is given by

$$y_t = \mu + \exp(\frac{1}{2}\theta_t)\varepsilon_t, \qquad \varepsilon_t \sim \mathrm{NID}(0, \sigma^2), \qquad t = 1, \ldots, n, \tag{5}$$

where μ is a constant and ε_t is normally distributed with zero mean and variance σ^2. The dynamic specification for θ_t can be determined by (2) and (3). The conditional observation density (1) is given by

$$p(y_t|\theta_t; \psi) = -\frac{1}{2}\log(2\pi\,\sigma^2) - \frac{1}{2}\theta_t - \frac{1}{2\sigma^2}\exp(-\theta_t)(y_t - \mu)^2, \qquad t = 1, \ldots, n.$$

The stochastic volatility model can be extended in many ways. For example, the Gaussian density for ε_t can be replaced by the Student's t density as in (4); we refer to Shephard (2005) for extensive discussions on stochastic volatility models.

Time-varying model for counts

Time series of counts can be modelled by the Poisson density with the intensity parameter as a function of the time-varying signal θ_t that we can specify as (2) and (3). The observation density is then given by

$$p(y_t|\theta_t;\psi) = y_t \log\theta_t - \theta_t - \log(y_t!), \qquad t = 1,\dots,n. \tag{6}$$

Other densities from the exponential family can also be considered such as the Binomial density. More examples are documented in Durbin and Koopman (2001, Chapter 10).

3 Efficient importance sampling

We discuss our implementation of the efficient importance sampling (EIS) method by considering likelihood evaluation. In the next section we discuss other applications in which EIS plays an important role including signal extraction, maximum likelihood estimation of ψ and forecasting of future observations y_t.

The likelihood function of the model (1), (2) and (3) for the observed vector $y = (y_1',\dots,y_n')'$ and as a function of parameter vector ψ can be represented by

$$L(y;\psi) = \int p(y,\alpha;\psi)\,\mathrm{d}\alpha = \int p(y|\alpha;\psi)p(\alpha;\psi)\,\mathrm{d}\alpha, \tag{7}$$

where $\alpha = (\alpha_1',\dots,\alpha_n')'$. This expression may suggest a basic Monte Carlo evaluation of the likelihood function via

$$\widehat{L}(y;\psi) = \sum_{i=1}^{M} p(y|\alpha^{(i)};\psi), \qquad \alpha^{(i)} \sim p(\alpha;\psi), \tag{8}$$

where $\alpha^{(i)}$ refers to the ith simulated sample of α that is generated from the unconditional density $p(\alpha;\psi)$ with $i = 1,\dots,M$. The standard law of large numbers (LLN) insists that $\widehat{L}(y;\psi)$ converges to $L(y;\psi)$ as $M \to \infty$. However, since the simulation of α has no reference to the data vector y, the efficiency of the estimate will be very low and therefore we need an extremely large value of M. Numerical evaluation of (7) is also not feasible given the high dimensional vector α.

Given the serial independence properties for the observations y_t conditional on θ_t and for the disturbances η_t, we have

$$L(y;\psi) = \int \left[\prod_{t=1}^{n} p(y_t|\alpha_t;\psi) p(\alpha_t|\alpha_{t-1};\psi) \right] \mathrm{d}\alpha, \tag{9}$$

with $p(\alpha_1|\alpha_0;\psi) = p(\alpha_1;\psi)$ and where $p(y_t|\alpha_t;\psi) = p(y_t|\theta_t;\psi)$ given the signal specification (2). We also have $p(\alpha_t|\alpha_{t-1};\psi) = p(\eta_{t-1};\psi) = \mathrm{NID}(0, Q_{t-1})$ for $t = 1, \ldots, n$.

3.1 Importance density

The numerical evaluation of (9) becomes feasible when using Monte Carlo methods and we consider the method of importance sampling; see Ripley (1987) for an introduction to simulation methods. For the purpose of evaluating (9) via importance sampling, we introduce the importance density based on the linear Gaussian joint density $g(y, \alpha; \psi)$ with properly defined mean vector and variance matrix. Its dependence on ψ is derivative from the original model (1) – (3). The decomposition $g(y, \alpha; \psi) = g(y|\alpha; \psi)\, g(\alpha; \psi)$ is valid and we assume that

$$g(y|\alpha;\psi) = \prod_{t=1}^{n} g(y_t|\theta_t;\psi), \qquad g(\alpha;\psi) = p(\alpha;\psi) = \prod_{t=1}^{n} p(\alpha_t|\alpha_{t-1};\psi),$$

since $g(y_t|\alpha_t;\psi) = g(y_t|\theta_t;\psi)$ from (2) and where $p(\alpha_t|\alpha_{t-1};\psi) = \mathrm{NID}(0, Q_{t-1})$, for $t = 1, \ldots, n$. Since the dynamic specification for the state vector in (3) is linear Gaussian, we adopt the same state specification for the importance density. The Gaussian observation density is expressed by

$$g(y_t|\theta_t;\psi) = \exp\left(a_t + b_t'\theta_t - \frac{1}{2}\theta_t' C_t \theta_t \right), \qquad t = 1, \ldots, n, \tag{10}$$

where $z_t = z_t(y_t;\psi)$ is a known function of y_t and ψ for $z = a, b, C$ and $t = 1, \ldots, n$. The constant a_t is needed to ensure that $g(y_t|\theta_t;\psi)$ integrates to one. The key functions are b_t and C_t which determine the mean and variance of the density $g(y_t|\theta_t;\psi)$. An effective importance sampler is obtained by selecting appropriate values for b_t and C_t for $t = 1, \ldots, n$. Hence the design of the importance sampler is elegantly reduced to a choice for b_t and C_t.

We notice that the importance density can also be expressed in terms of the constructed variable $x_t = C_t^{-1} b_t$ and the linear Gaussian model

$$x_t = \theta_t + u_t, \qquad u_t \sim \mathrm{NID}(0, C_t^{-1}), \qquad t = 1, \ldots, n, \tag{11}$$

where the Gaussian disturbances $u_1, \ldots, u_n$ are serially uncorrelated. We can show that the conditional obervation density function (10) is equivalent to $g(x_t|\theta_t;\psi)$; in logs, we have

$$\begin{aligned}\log g(x_t|\theta_t;\psi) &= -\frac{1}{2}\log 2\pi - \frac{1}{2}\log|C_t| - \frac{1}{2}\left\{(x_t-\theta_t)'C_t^{-1}(x_t-\theta_t)\right\} \\ &= a_t + b_t'\theta_t - \frac{1}{2}\theta_t'C_t\theta_t,\end{aligned}$$

since $x_t = C_t^{-1}b_t$, where a_t collects all terms that are not associated with θ_t. Hence it follows immediately that

$$g(y|\alpha;\psi) \equiv \prod_{t=1}^{n} g(x_t|\theta_t;\psi),$$

when we assume that x_t is modelled by (11) for $t = 1, \ldots, n$.

3.2 Likelihood evaluation via importance sampling

The actual importance density for the evaluation of (7) is chosen as

$$g(\alpha|y;\psi) = g(y|\theta;\psi)\, g(\alpha;\psi) \,/\, g(y;\psi),$$

where $g(\alpha;\psi) = p(\alpha;\psi)$. The likelihood function (7) with the importance sampling density incorporated is given by

$$L(y;\psi) = \int \frac{p(y|\alpha;\psi)p(\alpha;\psi)}{g(\alpha|y;\psi)} g(\alpha|y;\psi)\,\mathrm{d}\alpha.$$

After some minor manipulations, we can express the likelihood function as

$$L(y;\psi) = g(y;\psi)\int\left[\prod_{t=1}^{n} w(y_t,\alpha_t;\psi)\right] g(\alpha|y;\psi)\,\mathrm{d}\alpha, \qquad w(y_t,\alpha_t;\psi) = \frac{p(y_t|\alpha_t;\psi)}{g(y_t|\alpha_t;\psi)}, \qquad (12)$$

where $w(y_t,\alpha_t;\psi)$ is referred to as the importance weight, for $t = 1, \ldots, n$.

The evaluation of the likelihood function by means of importance sampling takes place by simulating state vectors from the importance density $g(\alpha|y;\psi)$ which we denote by

$$\alpha^{(i)} = \left(\alpha_1^{(i)\prime}, \ldots, \alpha_n^{(i)\prime}\right)' \sim g(\alpha|y;\psi), \qquad i = 1, \ldots, N.$$

Since we can represent $g(y,\alpha;\psi)$ by the linear Gaussian state space model (11), (2) and (3), we can simulate α from the smooth density $g(y,\alpha;\psi)$ via the simulation smoothing method; see, for example, Fruhwirth-Schnatter (1994), Carter and Kohn (1994), de Jong and Shephard (1995) and Durbin and Koopman (2002). Given the simulated realisations $\alpha^{(i)}$, for $i = 1, \ldots, N$, the likelihood function is computed by

$$\widehat{L}(y;\psi) = g(y;\psi) M^{-1} \sum_{i=1}^{N} \prod_{t=1}^{n} w_{it}, \qquad w_{it} = w(y_t, \alpha_t^{(i)};\psi). \tag{13}$$

Some practical guidance for the computation of $\widehat{L}(y;\psi)$ for the purpose of parameter estimation is given in Section 4.1. We expect that the Monte Carlo estimate (13) is more efficient than the estimate (8) since we simulate $\alpha_t^{(i)}$ with a reference to the data vector y.

3.3 Efficient importance sampling

The values for b_t and C_t, with $t = 1, \ldots, n$, need to be determined before the calculation of (13) can start. Here we follow Richard and Zhang (2007) and adopt their efficient importance sampling (EIS) method. They propose to choose b_t and C_t such that the criterion

$$I_t = \int \lambda_t^2(y_t, \alpha_t;\psi) p(y_t, \alpha_t;\psi)\, \mathrm{d}\alpha_t, \qquad \lambda_t(y_t, \alpha_t;\psi) = \log w(y_t, \theta_t;\psi) - \bar{\lambda}, \tag{14}$$

is minimized for each t separately and where $\bar{\lambda}$ is the normalizing constant such that the expectation of λ_t with respect to the true model $p(y_t, \alpha_t;\psi)$ is zero. We therefore interpret I_t as the variance of the logged importance weight function with respect to $p(y_t, \alpha_t;\psi)$. We notice that the variables b_t and C_t determine $g(y_t|\theta_t;\psi)$ that is part of $w(y_t, \theta_t;\psi)$ and hence of $\lambda_t(y_t, \alpha_t;\psi)$. The function I_t cannot be evaluated analytically for the same reason as the likelihood function (7) cannot be evaluated analytically. Hence we follow the same approach of introducing the importance density $g(\alpha_t|y;\psi)$. The criterion to be minimized can then be expressed as

$$\begin{aligned} I_t &= \int \lambda_t^2(y_t, \alpha_t;\psi) \frac{p(y_t, \alpha_t;\psi)}{g(\alpha_t|y;\psi)} g(\alpha_t|y;\psi)\, \mathrm{d}\alpha_t \\ &\propto \int \lambda_t^2(y_t, \alpha_t;\psi) \frac{p(y_t|\alpha_t;\psi)}{g(y_t|\alpha_t;\psi)} g(\alpha_t|y;\psi)\, \mathrm{d}\alpha_t \\ &\propto I_t^*, \end{aligned}$$

where

$$I_t^* = \int \lambda_t^2(y_t, \alpha_t;\psi) w(y_t, \theta_t;\psi) g(\alpha_t|y;\psi)\, \mathrm{d}\alpha_t. \tag{15}$$

The statements above are valid since b_t and C_t only have an impact on y_t and θ_t. Also, we have

$$g(\alpha_t|y;\psi) \propto g(y_t|\alpha_t;\psi)g(\alpha_t;\psi),$$

with $g(\alpha_t;\psi) = p(\alpha_t;\psi)$. The minimization of I_t^* with respect to (b_t, C_t) is therefore equivalent to the minimization of I_t. The evaluation and minimization of I_t^* takes place via importance sampling. We minimize

$$\widehat{I}_t^* = M^{-1}\sum_{i=1}^{M}\lambda_t^2(y_t,\alpha_t^{(i)};\psi)w(y_t,\theta_t^{(i)};\psi), \qquad \theta^{(i)} = Z_t(\alpha_t^{(i)}),$$

where $\alpha_t^{(i)}$ is obtained by sampling from $g(\alpha|y;\psi)$. This minimization of $\widehat{I}_t^*$ leads to the weighted least squares solution. In case θ_t is a scalar, we define the regression coefficient vector $\beta_t = (a_t^*, b_t, C_t)$ and the minimum is obtained at

$$\widehat{\beta}_t = \left(\sum_{i=1}^{M} w_{it}v_{it}v_{it}'\right)^{-1}\sum_{i=1}^{M} w_{it}v_{it}p_{it}, \tag{16}$$

where w_{it} is defined in (13) and where

$$v_{it} = (1, \theta_t^{(i)}, \theta_t^{(i)\,2})', \qquad p_{it} = \log p(y_t|\theta_t^{(i)};\psi),$$

for $t = 1, \ldots, n$. The sampling of $\alpha_t^{(i)}$ requires the simulation smoother that is applied to $g(y,\alpha;\psi) = g(y|\alpha;\psi)p(\alpha;\psi)$, in particular its model representation (11), (2) and (3). However, observation equation (11) requires values for b_t and C_t, for $t = 1, \ldots, n$, which we want to establish via the least squares solution (16). Since the Gaussian kernel of the log-density $\log g(y|\alpha;\psi)$ acts effectively as a second order Taylor approximation to $\log p(y|\alpha;\psi)$, around some value of θ_t, we can carry out the minimization iteratively as follows. We set values for $b_1, \ldots, b_n$ and $C_1, \ldots, C_n$ initially. A search for good starting values can be conducted but in many cases of practical interest, any set of initial values work sufficiently well. Next we simulate $\theta_t^{(i)}$ by means of simulation smoothing applied to the linear Gaussian model (11), (2) and (3) based on the current set of values for (b_t, C_t) for $t = 1, \ldots, n$. A new set of values can be obtained from (16). This iterative scheme continues until some level of convergence is obtained. It is assumed that at each iteration when samples are generated from $g(\theta|y;\psi)$ using a new set of values for (b_t, C_t), the same random numbers are used (or the same random seed is used) for computing $\alpha^{(i)}$ so that a smooth convergence process takes place.

3.4 A comparison with EIS

Our proposed implementation of the efficient importance sampling (EIS) method is clearly different than the one proposed by Richard and Zhang (2007) although the objective function is the same. The key insight that we explore is the representation of $g(y, \alpha; \psi)$ by the linear Gaussian state space model (11), (2) and (3) for the constructed variable x_t. This allows us to treat the EIS method on the basis of the computationally efficient Kalman filter and its related smoothing methods including the simulation smoother; see Durbin and Koopman (2001) for a treatment of state space methods.

Richard and Zhang (2007) have proposed the minimization of (14) and have provided the solution (16). The key difference is how the draws $\alpha_t^{(i)}$ are generated. In their implementation of EIS, they adopt an approximate backwards scheme, starting from $t = n$ towards $t = 1$, and need to track an integration constant so that each density at time t integrates to unity. We circumvent this time-consuming process since we interpret the density as a well-defined model for x_t and apply the simulation smoothing method of Durbin and Koopman (2002) for computing the draws $\theta_t^{(i)}$ directly.

Another key development in our implementation of the EIS is that the simulations in Richard and Zhang (2007) are with respect to the state vector α_t while the simulations in our implementation of EIS is based on the signal vector θ_t. In many empirical models of interest, the state vector is typically of a higher dimension than the signal vector which has the same dimension of y_t. We therefore expect that in many studies, our implementation will gain computational efficiency.

4 Nonlinear non-Gaussian state space analysis

In this section we briefly illustrate other applications of efficient importance sampling and provide the details for an effective implementation.

4.1 Maximum likelihood estimation of ψ

In Section 3.2 we have shown how the likelihood function can be evaluated by the method of importance sampling. The maximum likelihood estimate (MLE) of parameter vector ψ can be simply obtained via a numerical optimization method. Quasi-Newton methods are often used for this task. It may be clear that analytical expressions for the MLE are not available in almost all cases.

A number of numerical issues need to be addressed before the actual maximization of the likelihood function can take place. We evaluate the likelihood function as a Monte Carlo estimate. The use of different sets of random values for generating the importance draws of α^i, with $i = 1, \ldots, M$, leads clearly to different estimates of the likelihood function $L(y;\psi)$. Since numerical optimization methods require smooth functions, we evaluate the likelihood functions using the same set of random values. In other words, the same "seed" of the random number generator is taken for each likelihood evaluation. The likelihood is then a smooth function of ψ only.

In practice, the loglikelihood function is maximized. However, the log of the estimate (13) is not equal to the estimate of the loglikelihood function. The bias in the log of the estimate can be approximately corrected on the basis of a second-order Taylor expansion. We therefore maximize the bias-corrected loglikelihood estimate

$$\widehat{\ell(y;\psi)} = \log \widehat{L}(y;\psi) + \frac{1}{2M}\bar{w}^{-2}s_w^2, \qquad s_w^2 = (M-1)^{-1}\sum_{i=1}^{M}(w_i - \bar{w})^2,$$

where $\ell(y;\psi) = \log L(y;\psi)$, $w_i = \prod_{t=1}^{n} w_{it}$ and $\bar{w} = M^{-1}\sum_{i=1}^{M} w_i$; see Durbin and Koopman (1997) for more details.

The bias-corrected loglikelihood estimate can be expressed as

$$\widehat{\ell(y;\psi)} = \log g(y;\psi) + \log \bar{w} + \frac{1}{2M}\bar{w}^{-2}s_w^2, \tag{17}$$

The computation of w_i, $\log \bar{w}$ and $\bar{w}^{-2}s_w^2$ requires modifications for a numerically feasible and stable implementation. Define

$$a_i = \log w_i = \sum_{t=1}^{n} \log p(y_t|\alpha_t^{(i)};\psi) - \log g(y_t|\alpha_t^{(i)};\psi), \qquad \bar{a} = M^{-1}\sum_{j=1}^{M} a_j,$$

for $i = 1, \ldots, M$. The computation of a_i and $\bar{a}$ is numerically stable. However, the computation of $w_i = \exp(a_i)$ can lead to numerical overflow problems whereas the computation of $u_i = \exp(a_i - \bar{a})$ is numerical stable. It follows that $w_i = \exp(\bar{a})u_i$. After some further minor manipulations, it can be shown that

$$\log \bar{w} = \bar{a} + \log \bar{u}, \qquad \text{and} \qquad \bar{w}^{-2}s_w^2 = \bar{u}^{-2}s_u^2,$$

where

$$u_i = \exp(a_i - \bar{a}), \qquad \bar{u} = M^{-1}\sum_{i=1}^{M} u_i, \qquad s_u^2 = (M-1)^{-1}\sum_{i=1}^{M}(u_i - \bar{u})^2.$$

The bias-corrected loglikelihood estimate (17) is computed in a numerically feasible manner using these results.

4.2 Signal extraction : estimation of α_t and θ_t

The estimation of α_t is based on the evaluation of the integral

$$\tilde{\alpha} = \int \alpha p(\alpha|y;\psi)\,\mathrm{d}\alpha.$$

We have argued that also the evaluation of such integral in a computational efficient way can be carried out by efficient importance sampling. The construction of a Monte Carlo estimate for $\tilde{\alpha}$ is based on

$$\begin{aligned}\tilde{\alpha} &= \int \alpha\,[p(\alpha|y;\psi)\,/\,g(\alpha|y;\psi)]\,g(\alpha|y;\psi)\,\mathrm{d}\alpha \\ &= [g(y;\psi)\,/\,p(y:\psi)] \int \alpha w(y,\alpha;\psi) g(\alpha|y;\psi)\,\mathrm{d}\alpha, \end{aligned} \tag{18}$$

since $g(y|\alpha;\psi) = p(y|\alpha;\psi)$, where

$$w(y,\alpha;\psi) = \frac{p(y|\alpha;\psi)}{g(y|\alpha;\psi)} = \prod_{t=1}^{n} w(y_t,\alpha_t;\psi),$$

with $w(y_t,\alpha_t;\psi)$ as defined in (12). The density $p(y;\psi)$ reflects the likelihood function (12) and its substitution in (18) leads to the equation

$$\tilde{\alpha} = \frac{\int \alpha w(y,\alpha;\psi) g(\alpha|y;\psi)\,\mathrm{d}\alpha}{\int w(y,\alpha;\psi) g(\alpha|y;\psi)\,\mathrm{d}\alpha}.$$

The two integrals can be evaluated by Monte Carlo simulation. The estimate of $\tilde{\alpha}$ is then given by

$$\widehat{\tilde{\alpha}} = \frac{\sum_{i=1}^{M} \alpha^{(i)} w_i}{\sum_{i=1}^{M} w_i},$$

where $w_i = \prod_{t=1}^{n} w_{it}$ with w_{it} defined as in (13) and where both $\alpha^{(i)}$ and w_i are based on the draws from the importance density, that is

$$\alpha^{(i)} \sim g(\alpha|y;\psi), \qquad i = 1,\ldots,M.$$

The draws are obtained by using the method of efficient importance sampling described in Section 3.3. The nominator and denominator are typically computed by using the same random numbers and therefore we can base the estimate on normalised weights, that is

$$\widehat{\tilde{\alpha}} = \sum_{i=1}^{M} \alpha^{(i)} w_i^*, \qquad w_i^* = \frac{w_i}{\sum_{i=1}^{M} w_i}.$$

The signal is a function of the state vector, we have $\theta_t = Z_t(\alpha_t)$ and $\theta = Z(\alpha)$ where $Z(\alpha) = [Z_1(\alpha_1)', \ldots, Z_n(\alpha_n)']'$. Using the same arguments as above, the estimate of θ is given by

$$\tilde{\theta} = \frac{\int \theta w(y, \alpha; \psi) g(\alpha|y; \psi) \, \mathrm{d}\alpha}{\int w(y, \alpha; \psi) g(\alpha|y; \psi) \, \mathrm{d}\alpha},$$

and we evaluate it via the efficient importance sampling method to obtain

$$\widehat{\tilde{\theta}} = \sum_{i=1}^{M} Z(\alpha^{(i)}) w_i^*,$$

where the normalised weight w_i^* is defined above. These arguments are also valid for any other known function of α. State vectors at time periods after n can be estimated using this approach. It facilitates forecasting of state and signal vectors in time series models of this class.

5 Two Illustrations

5.1 Time-varying model for counts : some simulation evidence

To illustrate the EIS method, our modified EIS method and the SPDK method, we consider a simulation study for a time-varying model for counts with density (6), with time-varying signal θ_t specified by (2) and (3) and with state initialisation $\alpha_1 \sim N(0, \sigma_\eta^2/(1-\phi^2))$. We simulate time series of counts y_t for $t = 1, \ldots, n$ and $n = 1000$. The parameter vector is set equal to $(\phi, \log \sigma_\eta^2) = (0.9, -2)$. For each simulated time series, we estimate the two unknown parameter coefficients of ψ by maximizing the simulated likelihood function as defined by (13). For the numerical maximisation we adopt the BFGS optimization algorithm with starting values equal to the true parameter values. Likelihood evaluation is based on three methods SPDK, EIS and the modified EIS (MEIS). For the EIS and MEIS methods, we take $M = 50$ draws for constructing the importance density. Subsequently, we take $M = 500$ draws for the Monte Carlo likelihood evaluation. The empirical distributions of the estimates for the two parameters and for the three methods are presented in Figure 1. Although the empirical distributions are different from each other, for each parameter the distributions are centered around the true parameter values while the variations around the sample mean have similar sizes. However, the computer times for the three different methods are different. SPDK is fast as it fully relies on state space methods. The EIS method is slow since the construction of the importance density also relies on simulation. In the MEIS method, the importance density is obtained using state space methods and is therefore almost as fast as the SPDK method.

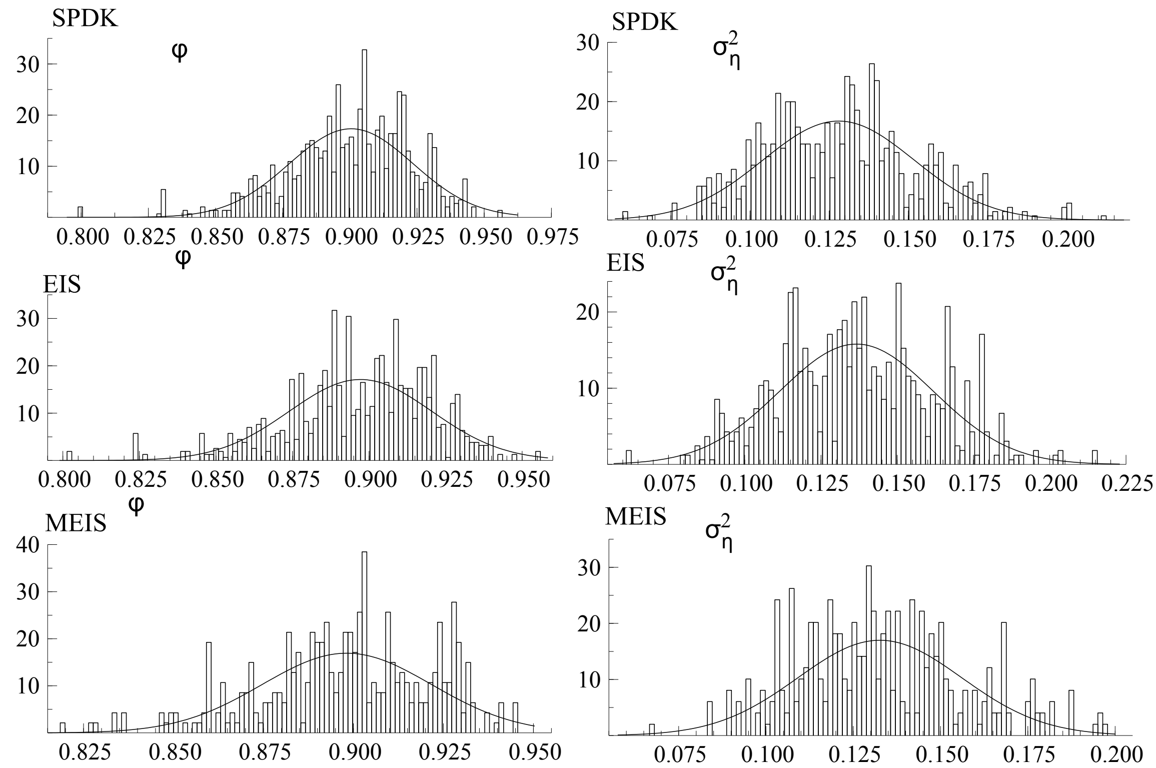

Figure 1: Empirical distributions for the two parameters of the time-varying model for counts as estimated by the three methods SPDK, EIS and the modified EIS (MEIS).

5.2 Stochastic volatility model for pound/dollar exchange rates

To empirically illustrate our modified EIS method, we analyse the volatility of log returns of pound/dollar exchange rates from 1-Oct-1981 to 28-Jun-1985 which is time series also analysed by Harvey, Ruiz, and Shephard (1994) and Durbin and Koopman (2001). The stochastic volatility model is specified as (5) with time-varying signal θ_t that is specified by (2) and (3) where the initial state is given by $\alpha_1 \sim N(0, \sigma_\eta^2/(1-\phi^2))$. Parameter estimation takes place using the modified EIS method. The first step is to find a suitable approximating linear Gaussian state space model to (5) with the observation density given by

$$\hat{y}_t = \theta_t + u_t, \qquad u_t \sim N\left(0, d_t^{-1}\right), \qquad t = 1, \ldots, n, \tag{19}$$

where $\hat{y}_t = b_t/d_t$ and where the coefficients b_t and d_t are obtained via applying least squares computations (16) repeatedly and are evaluated at the simulated samples $\theta^1, \ldots, \theta^M$ from the importance density $g(\theta|\hat{y})$ with $\hat{y} = (\hat{y}_1, \ldots, \hat{y}_n)'$. Simulation takes place via the simulation smoother of Durbin and Koopman (2002). The iteration process is initialised with the second-order Taylor coefficients evaluated at the mode. Once the coefficients have converged in the iteration process, we have obtained the importance density and the Monte Carlo estimate of the loglikelihood function can be evaluated.

The BFGS method is used to maximise the simulated loglikelihood function with respect to the parameter vector. The estimates of $\hat{\phi}$, $\hat{\sigma}_\eta^2$ and $\hat{\sigma}$ together with its standard errors are presented for the SPDK and MEIS methods in Table 5.2. We can conclude that the values of the estimated parameters obtained from the SPDK and MEIS methods are very close. The estimated volatility with a 95% confidence interval is displayed in Figure 2.

Parameter	SPDK	MEIS
ϕ	0.9731	0.9750
	(0.501)	(0.504)
σ_η	0.1726	0.1643
	(0.217)	(0.223)
σ	0.6338	0.6359
	(0.103)	(0.108)

Table 1: Simulated maximum likelihood parameter estimates for ϕ, σ_η, and σ using SPDK and MEIS methods based on $N = 500$ draws to estimate the likelihood function. The standard errors of the estimates are displayed between parantheses below.

References

Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 81*, 541–53.

Danielsson, J. and J. F. Richard (1993). Accelerated Gaussian importance sampler with application to dynamic latent variable models. *J. Applied Econometrics 8*, S153–S174.

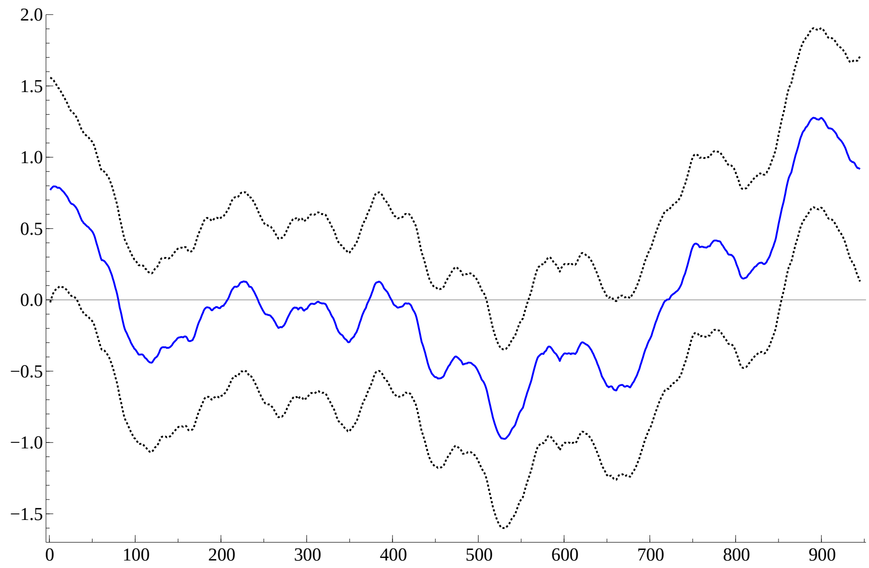

Figure 2: Estimated volatility process and the 95% confidence interval for the log returns of pound/dollar exchange rates from 1-Oct-1981 to 28-Jun-1985.

de Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 82*, 339–50.

Durbin, J. and S. J. Koopman (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika 84*, 669–84.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods.* Oxford: Oxford University Press.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*, 603–16.

Fruhwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *J. Time Series Analysis 15*, 183–202.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica 57*, 1317–39.

Harvey, A. C., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *Rev. Economic Studies 61*, 247–64.

Jungbacker, B. and S. J. Koopman (2007). Monte Carlo estimation for nonlinear non-Gaussian state space models. *Biometrika 94*, 827–39.

Kloek, T. and H. K. Van Dijk (1978). Bayesian estimates of equation system parameters:

an application of integration by monte carlo. *Econometrica 46*, 1–20.

Liesenfeld, R. and J. F. Richard (2003). Univariate and multivariate stochastic volatility models: Estimation and diagnostics. *J. Empirical Finance 10*, 505–531.

Richard, J. F. and W. Zhang (2007). Efficient high-dimensional importance sampling. *J. Econometrics 141*, 1385–1411.

Ripley, B. D. (1987). *Stochastic Simulation.* New York: Wiley.

Shephard, N. (2005). *Stochastic Volatility: Selected Readings.* Oxford: Oxford University Press.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*, 653–67.

So, M. K. P. (2003). Posterior mode estimation for nonlinear and non-Gaussian state space models. *Statistica Sinica 13*, 255–274.